

Automatische Aufgabengenerierung über Linked Open Data am Beispiel der Archäologie

Michael Striewe¹

Abstract: Automatische Aufgabengenerierung stellt fachspezifische Anforderungen an die verwendbaren Techniken. In manchen Fächern erscheint dabei die Nutzung von Linked Open Data vielversprechend. Dieses Paper untersucht am Beispiel der Archäologie die technischen und inhaltlichen Herausforderungen bei der Generierung von Multiple-Choice-Aufgaben. Die Untersuchung kommt zu dem Ergebnis, dass insbesondere die Vollständigkeit und inhaltliche Gewichtung der verfügbaren Daten ein Hindernis für die Realisierung darstellt.

Keywords: E-Assessment, Aufgabengenerierung, Linked Open Data, Semantic Web

1 Einleitung

Eine wiederkehrende Herausforderung in der Übungs- und Prüfungsvorbereitung ist die Erzeugung hinreichend vieler Aufgaben. Gerade im Übungsbetrieb ist es wünschenswert, ein Thema in voller Breite abdecken und viele Aufgaben auf unterschiedlichen Kompetenzniveaus anbieten zu können. Einen Lösungsansatz bieten hier Verfahren zur automatischen Aufgabenerzeugung [FR13, GLZ18], die in E-Assessments ggf. auch zur dynamischen Generierung unmittelbar beim Aufruf einer Aufgabe nutzbar sind. Im Fach Mathematik ist es bspw. gängig, Zahlen auszutauschen um beliebig viele Instanzen einer Aufgabe zu erzeugen [Ma18]. In textlastigen Fächern können Verständnisfragen zu Texten automatisch erzeugt werden [IGY13, MJ12]. Auf verschiedene Geistes- und Naturwissenschaften treffen beide Charakterisierungen jedoch nicht vollständig zu, wenn neben Zahlen, Formeln und Texten auch Abbildungen aller Art oder reale Gegenstände zum Einsatz kommen und Kompetenzen wie die sichere Beherrschung von Faktenwissen sowie die Anwendung von Kategorien und Konzepten geprüft werden sollen. Genau dies ist auch für die hier exemplarisch betrachtete Archäologie der Fall.

Die maschinenlesbare Abbildung von Fakten, Konzepten und Klassifikationen ist der Kern des „Semantic Web“, indem Informationen so mit Metadaten annotiert werden, dass sie verknüpft und ausgewertet werden können [SBH06]. Grundlage dafür sind Ontologien, die Entitäten und ihre möglichen Eigenschaften und Verknüpfungen beschreiben. Es existieren sowohl freie und thematisch offene derartige Datenbanken², als auch fachspezifische Datenbanken [La15]. Diese Datensammlungen mit Querverweisen untereinander werden auch als „Linked (Open) Data“ bezeichnet [Bi08].

¹ Universität Duisburg-Essen, paluno - The Ruhr Institute for Software Technology, Gerlingstraße 16, 45127 Essen, michael.striewe@paluno.uni-due.de

² Beispielsweise <https://www.wikidata.org/> oder <https://wiki.dbpedia.org/>

Angesichts der freien Verfügbarkeit von Daten und standardisierten Technologien stellt sich die Frage, ob es möglich ist, sinnvolle Aufgaben für die Archäologie automatisiert zu erzeugen. Unbeschadet der theoretischen Machbarkeit wird die praktische Realisierbarkeit, aus technischer Sicht davon beeinflusst, ob Daten schnell genug verfügbar sind und aus fachlicher Sicht dadurch, ob die verfügbaren Daten korrekt sind, hinreichend vollständig um anspruchsvolle Aufgaben zu erstellen, und fachlich ausgewogen genug um ein Themengebiet in der nötigen Breite und Tiefe abzudecken. Letztlich steht hinter diesen Fragen die Evaluation der Nutzbarkeit von Linked Open Data Ressourcen zum Zwecke des E-Assessment. Da sich der Datenbestand in editierbaren Datenbanken naturgemäß jederzeit ändern kann, sind wertende Aussagen jedoch immer nur für einen gegebenen Zeitpunkt möglich. Alle Aussagen in diesem Paper beziehen sich daher auf einen Untersuchungszeitraum von Mitte Januar bis Mitte Februar 2019.

1.1 Verwandte Arbeiten

Arbeiten zur Aufgabengenerierung auf Basis von Linked (Open) Data bzw. Ontologien zeigen die grundsätzliche Umsetzbarkeit des Ansatzes [A11, Sh19]. Die konkrete Nutzung offener, allgemeiner Datenbanken wurde ebenfalls bereits untersucht [Fo12, St18], ohne jedoch eine fachspezifische Evaluation durchzuführen. Allgemeine Ansätze, um die Qualität von Linked Open Data zu bewerten [Za16] oder systematisch zu verbessern [Ha18] gibt es ebenso wie Ansätze, die Qualität von offenen Daten in Relation zu einem bestimmten Einsatzzweck zu untersuchen [MZF17]. Untersuchungen mit Fokus auf die Aufgabengenerierung sind jedoch nicht verbreitet.

1.2 Technischer Hintergrund

Ein standardisiertes Format für Linked Data ist das „Resource Description Framework“ (RDF)³. Die semantische Beschreibung von Daten basiert auf Tripeln der Form „Subjekt – Prädikat – Objekt“, wobei Subjekt und Objekt Entitäten sind, während das Prädikat einer semantischen Beziehung in der zugrundeliegenden Ontologie entspricht. Das Subjekt könnte bspw. ein archäologisches Fundstück sein und das Objekt ein geographischer Ort. Ein Prädikat könnte dieser Beziehung dann die Bedeutung „Fundort“ zuweisen, ein anderes Prädikat dagegen die Bedeutung „Aufbewahrungsort“. Abfragen auf derartigen Daten können die ebenfalls standardisierte Abfragesprache SPARQL verwenden. Mit dieser können als Fortführung des o.g. Beispiels über eine Abfrage alle Objekte gefunden werden, für die derselbe Fundort oder Aufbewahrungsort hinterlegt ist. Über Mengenoperationen, Unterabfragen und Filter können auch komplexere Zusammenhänge zwischen Datenbeständen ausgewertet werden.

Die Umsetzung in diesem Paper basiert auf dem E-Assessment-System JACK, in dem Autoren Aufgaben mit Platzhaltern erstellen können, die dynamisch belegt werden. Die

³ <https://www.w3.org/RDF/>

dazu benötigten Belegungsvorschriften können auch in Form von SPARQL-Abfragen angegeben werden, die beim Laden einer Aufgabe ausgeführt werden [St18].

2 Fallbeispiel 1: Historische Geographie

In der historischen Geographie lernen Studierende antike Städte, archäologische Stätten sowie historische Landschaftsbezeichnungen kennen und müssen sich dazu eine große Menge Faktenwissen aneignen, um bspw. im weiteren Studienverlauf in der provinzial-römischen Archäologie in der Lage zu sein, alle römischen Provinzen sowie wichtige Städte und bekannte Fundstätten in diesen Provinzen namentlich benennen und korrekt zuordnen zu können. Entsprechende Übungen sind daher Teil der Einführungsphase des Studiums. Eine einfache Umsetzung einer passenden Aufgabe benennt eine römische Provinz und fragte im Multiple-Choice-Verfahren nach der Zugehörigkeit von vier Städten. Von diesen kann eine oder mehrere korrekt sein.⁴

Zur Abfrage der benötigten Daten von Wikidata kann eine SPARQL-Anfrage (Abb. 1) verwendet werden, die im SELECT-Teil definiert, dass Paare von Orten und Provinzen gesucht werden. Im WHERE-Teil fragt sie zunächst Entitäten ab, die zur Kategorie Stadt (Q515) oder archäologische Stätte (Q839954) gehören. Ferner muss für diese Entitäten eine Zuordnung zu einem administrativen Gebiet vorliegt (P131) und bei diesem muss es sich um eine antike römische Provinz (Q182547) handeln. Filter begrenzen das Ergebnis auf Orte und Provinzen, für die die deutsche Bezeichnung vorliegt.

```
SELECT DISTINCT ?cityLabel ?provLabel WHERE {
  {?city wdt:P31 wd:Q515.} UNION {?city wdt:P31 wd:Q839954.}
  ?city wdt:P131 ?prov. ?prov wdt:P31 wd:Q182547.
  ?prov rdfs:label ?provL. ?city rdfs:label ?cityL.
  FILTER (langMatches( lang(?provL), "de" ) )
  FILTER (langMatches( lang(?cityL), "de" ) )
  SERVICE wikibase:label {bd:serviceParam wikibase:language "de".}
}
```

Abb. 1: SPARQL-Anfrage zur Abfrage von Städten und archäologischen Stätten in römischen Provinzen aus der Wikidata-Datenbank.

Aus dem Ergebnis wählt JACK bei Aufruf der Aufgabe zufällig vier Paare aus. Die Provinz aus dem ersten Paar wird im Aufgabentext angegeben. Die vier Orte aus allen Paaren bilden in zufälliger Reihenfolge die Antwortoptionen.

Die Laufzeit der Abfrage beträgt weniger als eine Sekunde und ist damit hinreichend schnell für die dynamische Aufgabengenerierung im Live-Betrieb. Alle erhaltenen Paare von Orten und Provinzen sind inhaltlich korrekt. Insgesamt liefert die Abfrage Paare mit 55 Orten in 16 Provinzen, was eine überraschend geringe Zahl darstellt, da tatsächlich deutlich mehr Orte und römische Provinzen in Wikidata erfasst sind. Grund ist, dass die Eigenschaft P131 in Verbindung mit einer römischen Provinz für die meisten Orte nicht

⁴ Eine Beispielaufgabe ist verfügbar unter <https://jack-demo.s3.uni-due.de/jack2/demo?id=63144>

gesetzt ist. Die verfügbaren Daten sind demnach deutlich unvollständig. Kombinatorisch lassen sich aus den 55 Paaren gut 340.000 individuelle Aufgaben erzeugen. Der Wert dieser Zahl wird jedoch durch einen ungünstigen inhaltlichen Schwerpunkt gemindert: Von den 55 Paaren beziehen sich 37 auf die Provinz „Germania inferior“, so dass mehr als 2/3 der Aufgaben diese Provinz im Aufgabentext nennen werden. Die Breite des Themas wird so nicht angemessen reflektiert. Zufriedenstellend lösbar sind beide Probleme, indem fehlende Informationen zur Verknüpfung von Orten und Provinzen in Wikidata nachgetragen werden, um Umfang und Vollständigkeit der Daten zu erhöhen.

Eine andere Möglichkeit, Zuordnungen zwischen Orten und Provinzen zu ergänzen besteht darin, diese Information automatisch aus geographischen Daten abzuleiten. Koordinaten für Orte stehen dazu über verschiedene Quellen zur Verfügung. Vollständige und genaue antike Grenzverläufe sind jedoch in keiner allgemeinen oder fachspezifischen Quelle verfügbar, so dass dieser Ansatz nicht realisierbar ist.

3 Fallbeispiel 2: Bestimmungsübung

Neben Faktenwissen gehört die Bestimmung von Objekten hinsichtlich fachlicher Klassifikationen zum grundlegenden Handwerkszeug eines Archäologen. Übungen, in denen Studierende bspw. Keramikgefäße aufgrund ihrer Form und Bemalung in fachliche Kategorien einordnen müssen, gehören daher ebenfalls zur Grundausbildung in der Archäologie. Die dazu benötigten Daten sind etwas komplexer als im ersten Fallbeispiel, da die Beziehung zwischen einem Objekt und mehreren Klassifikationen bekannt sein muss und die zu bestimmenden Objekte zudem als Bild vorliegen müssen. Eine beispielhafte Umsetzung⁵ einer passenden Aufgabe kann identisch zum ersten Fallbeispiel erfolgen: Es kommt wieder das Multiple-Choice-Verfahren zum Einsatz und es wird im Aufgabentext eine Klassifikation genannt, zu der eine oder mehrere der diesmal als Bild gegebenen Antwortoptionen passen. Die Aufgabe in einer Lernzieltaxonomie (z. B. [Kr02]) allerdings höher einzustufen als das erste Fallbeispiel, da Konzeptwissen angewandt werden muss, um die Aufgabe richtig zu lösen.

Da weder das benötigte Bildmaterial noch die fachlichen Klassifikationen über Wikidata zur Verfügung stehen, erfolgt die Realisierung über die Abfrage einer Fachdatenbank⁶. Die dabei verwendete SPARQL-Abfrage (Abb. 2) ist eine sog. Federated Query, die drei Datenquellen miteinander verknüpft. Im SELECT-Teil werden die gesuchten 4-Tupel (Objekt-ID, Bildreferenz, zwei Klassifikationen) eingeführt. Aus der Kerameikos-Datenbank werden dann im WHERE-Teil Objekte mitsamt der zugehörigen Abbildung (depiction) sowie ihre Klassifikation bezüglich Form (hasShape) und Bemalung (P32_used_general_technique) abgefragt. Die beiden letzten Informationen liegen in dieser Datenbank jedoch nur in Form einer Entitäten-URI vor, die über eine andere Datenbank in einen Namen aufgelöst werden muss. Dabei sind nicht alle Objekte in der

⁵ Eine Beispielaufgabe ist verfügbar unter <https://jack-demo.s3.uni-due.de/jack2/demo?id=68081>

⁶ <http://www.kerameikos.org/>

Kerameikos-Datenbank über Begriffe aus derselben Datenbank klassifiziert, weshalb der WHERE-Teil aus einer Vereinigung der Ergebnismenge aus zwei Unterabfragen besteht. Die erste Unterabfrage nutzt die Datenbank des British Museum zur Auflösung der Referenzen, die zweite die Datenbank des J. Paul Getty Trusts⁷. Die zusammengeführte Ausgabe wird abschließend zufällig sortiert und auf 20 Elemente begrenzt.

```
SELECT ?o ?p ?sl ?tl WHERE {
  { ?o kon:hasShape ?s. ?o foaf:depiction ?p. ?o crm:P32_used_general_technique ?t.
    SERVICE <https://collection.britishmuseum.org/sparql> {
      ?s skos:prefLabel ?sl. ?t skos:prefLabel ?tl.
    } } UNION
  { ?o kon:hasShape ?s. ?o foaf:depiction ?p. ?o crm:P32_used_general_technique ?t.
    SERVICE <http://vocab.getty.edu/sparql> {
      ?s xl:prefLabel ?slf. ?slf xl:literalForm ?sl.
      ?t xl:prefLabel ?tlf. ?tlf xl:literalForm ?tl.
      FILTER (lang(?sl) = "en") FILTER (lang(?tl) = "en")
    } } ORDER BY RAND() LIMIT 20
```

Abb. 2: SPARQL-Anfrage zur Abfrage von antiken Keramikobjekten und ihrer Klassifikation aus der Kerameikos-Datenbank.

Die Laufzeit der Abfrage beträgt über 45 Sekunden, was die Live-Generierung von Aufgaben träge erscheinen lässt. Die abgefragten Daten sind fachlich korrekt, aber unvollständig: Von den ca. 320 Objekten ist für 80 kein Bild hinterlegt, so dass diese für die Aufgabe nicht verwendbar sind. Zwei der Bildverweise sind zudem fehlerhaft, so dass sporadisch lückenhafte Aufgaben erzeugt werden. Ebenfalls problematisch ist, dass die zwei verschiedenen Klassifikationen der beiden zusätzlichen Datenbanken nicht automatisch miteinander abgeglichen werden können. Antwortoptionen können daher von JACK nicht zuverlässig als korrekt erkannt werden. Derartige Fehler können ausgeschlossen werden, wenn nur die Objekte eines Museums verwendet werden. So wird jedoch der Vorteil der Verfügbarkeit vernetzter Daten faktisch aufgehoben.

4 Bewertung und Fazit

Die in diesem Paper durchgeführte Analyse kommt auf der praktischen Ebene zu einem negativen Fazit: Während die Qualität der verfügbaren Daten hinreichend gut ist, stehen in den untersuchten Fällen nicht ausreichend viele bzw. nicht hinreichend vollständige Daten zur Verfügung, um das jeweilige Thema über generierte Aufgaben sinnvoll abdecken zu können. Diese Aussage ist jedoch spezifisch für die gewählten Beispiele, da für andere fachliche Bereiche vollständigere Daten zur Verfügung stehen könnten. Eine genaue Analyse für das jeweilige Thema ist daher unerlässlich. Trotzdem kann geschlossen werden, dass sich vor allem Umfang und Vollständigkeit der verfügbaren Linked Open Data Ressourcen noch verbessern müssen, bevor diese uneingeschränkt für die automatische Aufgabengenerierung in der Archäologie eingesetzt werden können.

⁷ <https://collection.britishmuseum.org/resource/sparql> und <http://vocab.getty.edu/sparql>

Literaturverzeichnis

- [Al11] Al-Yahya, M.: OntoQue: A Question Generation Engine for Educational Assessment Based on Domain Ontologies. In: 11th IEEE International Conference on Advanced Learning Technologies, 393-395 (2011).
- [Bi08] Bizer, C. et al.: Linked Data on the Web (LDOW2008). Proceedings of the 17th international conference on World Wide Web, 1265-1266, 2008.
- [Fo12] Foulonneau, M.: Generating Educational Assessment Items from Linked Open Data: The Case of DBpedia. In: The Semantic Web: ESWC 2011 Workshops, 16-27, 2012.
- [FR13] Foulonneau, M.; Ras, E.: Assessment Item Generation, the way forward. In: CAA 2013 International Conference, 2013.
- [GLZ18] Gierl, M., Lai, H., Zhang, X.: Automatic Item Generation. In Encyclopedia of Information Science and Technology, 4th Edition (pp. 2369-2379). IGI Global, 2018.
- [Ha18] Hadhiatma, A.: Improving data quality in the linked open data: a survey. In: Journal of Physics: Conference Series, 978, 2018.
- [IGY13] Iwane, N., Gao, C., Yoshida, M.: Question Generation for Learner Centered Learning. In: 13th IEEE Int. Conference on Advanced Learning Technologies, 330-332, 2013.
- [Kr02] Krathwohl, D.: A Revision of Bloom's Taxonomy: An Overview. In: Theory Into Practice, 41:4, 212-218, 2002.
- [La15] Langner, M.: Archäologische Datenbanken als Virtuelle Museen. In: Digital Classics Online 1, 2015, 46-70, <https://journals.ub.uni-heidelberg.de/index.php/dco/article/view/20314/14271>
- [MJ12] Mostow, J.; Jang, H.: Generating Diagnostic Multiple Choice Comprehension Cloze Questions. In: Proc. Workshop on Innovative Use of NLP for Building Educational Applications, 136–146, 2012.
- [Ma18] Massing, T. et al.: E-Assessment Using Variable-Content Exercises in Mathematical Statistics. In: Journal of Statistics Education, 26(3), 174-189, 2018.
- [MZF17] Mocnik, F.-B.; Zipf, A.; Fan, H.: Data Quality and Fitness for Purpose. In: 20th AGILE Conference on Geographic Information Science, 2017.
- [SBH06] Shadbolt, N.; Berners-Lee, T.; Hall, W.: The Semantic Web Revisited. In: IEEE Intelligent Systems 21(3): 96-101, 2006.
- [Sh19] Shi, W. et al.: A Framework for Automatically Generating Medical Quizzes with Multi-media Contents Based on Linked Data. In: Advances on Broadband and Wireless Computing, Communication and Applications, 147-158, 2019.
- [St18] Striewe, M.: Dynamic Generation of Assessment Items Using Wikidata. In: Proc. of the 21st Int. Technology Enhanced Assessment Conference (TEA 2018), im Druck.
- [Za16] Zaveri, A. et al.: Quality assessment for Linked Data: A Survey. In: Semantic Web, 7(1), 63-93, 2016.